Effectiveness of discriminant analysis and varimax rotation as food forensic tools for authentication of skin gelatine sources

Muhamad Shirwan Abdullah Sani (shirwansany@iium.edu.my)
Azilawati Mohd Ismail
Azman Azid
Mohd Saiful Samsudin

Research Article

Keywords: Amino acids, skin gelatine, UHPLC-DAD, discriminant analysis, principal component analysis

Posted Date: May 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1633547/v1

Abstract

This study aims at (1) authenticating sources of skin gelatine via the incorporation of putative amino acid (AA) analysis via Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC-DAD) with multivariate data analysis and (2) developing the amino acid profiles in skin gelatines. The classification ability of MDA, such as partial least square-discriminant analysis (PLS-DA) and discriminant analysis (DA) was compared to choose the best discriminating model. Principal component analysis (PCA) with Varimax rotation was executed to ensure the correct grouping of the skin gelatine clusters and thus, facilitate assigning significantly contributing AA to the skin gelatine clusters. The DA was superior over the PLS-DA since it had successfully classified 96.7% porcine, bovine and fish skin gelatines using 17 AAs. The backbone of chemical structure may render the correlations among AAs in the skin gelatines. The DA identified 15 significant AAs ($p < 0.01$) in the skin gelatines via PCA with Varimax rotation. Four Varimax rotations had successfully grouped the porcine, bovine and fish skin gelatines into correct clusters. L-Tyrosine, L-Phenylalanine and L-Valine were dominant in porcine gelatine; L-Methionine, L-Threonine, L-Serine, L-Histidine, L-Arginine and Glycine were dominant in fish gelatine, while L-Proline, L-Leucine and L-Hydroxyproline were in moderate content in bovine. This study anticipated that the authority might adopt this approach to establish an authentication standard for skin gelatine samples.

1.0 Introduction

Gelatine manufacturing worth USD1.34 billion utilised approximate 80% of hydrolysed skin gelatine to produce food gelatine products, e.g. marshmallows, gummies and candies, etc. (Yap & Gam, 2019). Due to a considerable revenue, gelatine manufacturers tend to taint food integrity (Soon et al., 2017) by declaring false claims on the skin gelatine sources and exposing consumers to mad cow or bovine spongiform encephalopathy disease from the consumption of bovine-source food (Azilawati et al., 2015). Furthermore, this false claim affects the confidence of Muslims (Department of Standards Malaysia, 2019), vegetarians (Mutalib et al., 2015) and Jews (Romi Mukherjee, 2014) consumers on the consumption of gelatine products. Hence, our study proposes an authentication approach for skin gelatine sources, i.e., porcine, bovine, and fish, incorporating Ultra-High-Performance Liquid Chromatography-Diode-Array Detector (UHPLC-DAD) and multivariate data analysis (MDA).

Besides UHPLC-DAD, application of Liquid Chromatography-Mass Spectrometer (LC/MS) and Liquid Chromatography Time-of-Flight Mass Spectrometer (LC-QTOF/MS) have also assisted the authentication of skin gelatine sources; however, these methods that employ mass spectrometer are costly for maintenance, albeit of their high sensitivity (Rohman et al., 2020). Nonetheless, Azilawati et al. (2015) adopted an affordable High-Performance Liquid Chromatography-Fluorescence Detector (HPLC-FLD) to measure amino acids (AAs) and coupled the analytical method with principal component analysis (PCA). Hence, our study adopted this cheaper method, except that our study utilised the diode array detector instead, which has been proven to authenticate the skin gelatine sources (Abdullah Sani et al., 2021).

Previous studies, including established testing methods, validate and verify analytical methods by establishing calibration curve, linearity range significant different of the intercept of the calibration curve from zero (Sani et al., 2020) prior to employing the MDA, e.g., principal component analysis, cluster analysis, discriminant analysis, etc. While these steps carried out by Ismail et al. (2021) and Abdullah Sani et al. (2021) incur the cost and extend the time to achieve the acceptable performance characteristics, Idris et al. (2021) proposed that the application of putative biomarkers that could serve as an alternative and quicker approach for halal authentication, especially in the case of ad-hoc analysis. Furthermore, negligible reports employed the putative approach to authenticate skin gelatine sources. Previous application of MDA also employed significant level (α) at 0.05 that allows possible 5% error. However, since the contamination of food with non-halal ingredients is a sensitive issue to the Muslim consumers; hence, authenticate the gelatine at α of 0.01 would bring more confidence to the results.

The MDA for authentication for food is mostly involving discriminant analysis (DA) and partial least square (PLS-DA) since they are the common MDA classification approach (Sharin et al., 2021). Abdullah Sani et al. (2021) authenticated the gelatine products using DA without comparing the DA classification ability with PLS-DA, which the information is still negligible. Besides, previous studies did not explore the DA, and PLS-DA features such as Fisher statistics, p-value and variable importance in the projection (VIP) that able to identify the significantly contributing amino acids to the classification functions. The lacking of employing these feautes may hinder the progress of the authentcaitcaion method for skin gelatine sources compared to other authentciation methods since these features have proven to reduce the dimentionality of fatty acids methyl ester and thus assisted authentciaion of oil-based products (Idris et al., 2021). Due to these reasons, our study compared the classification ability of DA and PLS-DA by exploring their features on the skin gelatine dataset.

Another approach to study the distribution of AAs in each source of skin gelatine is by using exploratory principal component analysis (PCA). Azilawati et al. (2015) and Ismail et al. (2021) carried out the PCA and assigned the contributing AAs by manually removing the skin gelatine samples from the dataset. Manually removing samples is undesirable since adopting this approach become a challenge to the testing laboratory, especially in obtaining repeatability results due to unstandardised procedure. Otto (2017) proposed conducting PCA with Varimax

rotation to achieve correctly clustered samples, and our study adopted this approach with various values of Varimax rotation until the skin gelatines were grouped into specific clusters. We could then assign the contributing AAs to the specific clusters upon carrying out this approach and develop the AA profiles in each cluster of the skin gelatine. Since this approach has not been adopted and explained in the previous studies, this approach is expected to be used by certification or regulatory authorities to produce guidelines or standards for skin gelatine authentication.

2.0 Methodology

2.1 Skin gelatine preparation

Porcine (G6144), bovine (G9382) and cold-water fish skin (G7041) gelatines of Sigma Aldrich, USA, were freeze-dried to achieve less than 10% moisture content. An approximate 0.2 g of 40 porcine, 40 bovine and 40 fish skin gelatines were acid-hydrolysed by mixing with 5 mL of 6 N hydrochloric acid and incubated at 110°C for 24 h. The hydrolysate solution was mixed with 100 pmol/µL Aaba, diluted to 100 mL with distilled water and filtered with 0.45 µm cellulose acetate membrane to produce mixture I.

2.2 Amino acids analysis of skin gelatines

A volume of 10 µL mixture I was derivatised at pH 8.2–10.0 by mixing it with 70 µL AccQ.Fluor borate buffer (Waters, Massachusetts, USA) and 20 µL AccQ.Fluor reagent (Waters, Massachusetts, USA) to produce mixture II. The mixture II was heated at 55°C for 10 min and spiked with 100 pmol/µL internal standard solution of L-Aminobutyric acid (Aaba) (Waters, the USA) to produce mixture III. A volume of 1 µL mixture III was injected into Ultra-High-Performance Liquid Chromatography Diode-Array Detector (UHPLC-DAD) of Agilent, USA, and eluted by pre-filtered eluant A (AccQ.Tag™ concentrate (WAT052890)) and B (acetonitrile). The elution was established via a gradient elution set-up by Ismail et al. 2021). The AA peaks were separated at 1 mL/min by a Waters AccQ.Tag column (3.9 mm x 150 mm) at 36°C and detected by a diode array detector (DAD) at 260 nm.

The detected AA peaks were confirmed by comparing their retention time with injected mixture of 17 standard solutions (SS) of AA hydrolysate (Waters, the USA) and Aaba. The SSS contained L-Histidine (His), L-Serine (Ser), L-Arginine (Arg), Glycine (Gly), L-Aspartic acid (Asp), L-Glutamic acid (Glu), L-Threonine (Thr), L-Alanine (Ala), L-Proline (Pro), L-Lysine (Lys), L-Tyrosine (Tyr), L-Methionine (Met), L-Valine (Val), L-Isoleucine (Ile), L-Leucine (Leu) and L-Phenylalanine (Phe), L-Cystine (Cys) and L-Hydroxyproline (Hyp). The AA concentration in each skin gelatine was presented in percentage, while for MDA, the ratio of peak area of SS over Aaba was computed for each gelatine and subjected to dataset pre-processing.

2.3 Dataset pre-processing

Three dataset types were prepared; training, cross-validation and testing datasets, and these datasets were pre-processed using XLSTAT-Pro (2019) statistical tools (Addinsoft, Paris, France). This pre-processing step involved outlier treatment, dataset transformation, and dataset adequacy test at a significant level (α) of 0.01.

The training dataset consisting 40 porcine, 40 bovine and 40 fish skin gelatines, which brought 120 skin gelatines x 17 AAs was prepared for the MDA. A dataset was also prepared to validate the discriminating ability of the partial least square-discriminant analysis (PLS-DA) and discriminant analysis (DA) models where 40 skin gelatines x 17 AAs were randomly selected by the XLSTAT-Pro (2019) statistical tools. Using an 80-to-20 minimum ratio of a training-to-testing dataset (Andrada et al., 2015), a testing dataset entailing 10 porcine, 10 bovine and 10 fish skin gelatines (30 skin gelatines x 17 AAs) was prepared for PLS-DA and DA.

Outlier identification was carried out using Grubbs and Dixon tests, and the detected outliers were replaced with the mean value of individual AA. To ensure normal distribution is followed, each AA normality was subjected to dataset transformation using standardise (n-1) method (Abdullah Sani et al., 2021).

The dataset adequacy test was carried out using Kaiser-Meyer-Olkin (KMO) test and the average KMO value was evaluated according to these ranks: KMO < 0.5 = inadequate, 0.5 < KMO < 0.7 = mediocre, 0.7 < KMO < 0.8 = good, 0.8 < KMO < 0.9 = very good and KMO > 0.9 excellent to indicate the dataset adequacy. Among these ranks, the KMO > 0.5 was acceptable for MDA (Ismail et al., 2021).

Post dataset pre-processing, the dataset was subjected to PLS-DA and varimax rotation of PCA.

2.4 Partial least square-discriminant analysis

This study employed PLS-DA to develop a discriminating model (DM) for the skin gelatine sources. A new column labelled as 'cluster' was added to the training, cross-validation and testing datasets and each gelatine was assigned as 'porcine', 'bovine' and 'fish' clusters. Likewise, the cross-validation dataset was subjected to outlier treatment and dataset transformation, while the testing dataset was transformed prior to

the PLS-DA. The PLS-DA was carried out twice; the first PLS-DA involved 17 AAs, and AAs with variable importance in the projection (VIP) score > 0.8 were selected and subjected to the second PLS-DA.

The PLS-DA was carried out at α of 0.01 (Saiful et al., 2019) on the training dataset to establish a discriminating model (DM) for the porcine, bovine and fish skin gelatines via this formula:

$$F\left(y_i, a_k\right) = b_0 + \sum_{j=1}^{p} b_i x_{ij}$$

Where F is the function, k is number of skin gelatine classes for y dependant variable, a is the skin gelatine class, b_0 is the DM's intercept, p is the number of AAs, b_i is DM's coefficient and x is the observation. The DM discriminates skin gelatines into a class via this formula:

$$x = argmax_k F(y_i, a_k)$$

The DM returns the argument (*arg*) of the F to the maximum (*max*) class, i.e. *k = 3* in this study.

The predictive ability of the DM was evaluated on the percentage of correct classification of the porcine, bovine and fish skin gelatines. Cluster dissimilarity of the skin gelatines was also evaluated via the Mahalanobis distance value and their p-value (Brereton & Lloyd, 2014). To validate the high predictive ability of the DM, the established DM was validated and tested on cross-validation, and testing datasets, respectively and their percentage of correct classifications were evaluated.

The AAs with VIP scores > 0.8 were identified and subjected to the second PLS-DA. The correct classification of the porcine, bovine, and fish skin gelatines was evaluated. Classification ability between the first and second PLS-DA model was compared, and the best PLS-DA model was selected.

2.5 Discriminant analysis

This study employed discriminant analysis (DA) to develop another DM for the gelatine sources by using the same dataset in PLS-DA consisting of 17 AAs. A new column labelled as' cluster' was also created, and the skin gelatines were assigned to 'porcine', 'bovine' and 'fish' clusters. The DA was executed at α of 0.01 (Saiful et al., 2019), namely as DAAA via this formula:

$C_1 = c_0 + b_1 a_1 + b_2 a_2 + ... b_i a_i$; and

$$C_2 = c_0 + b_1 a_1 + b_2 a_2 + ... b_i a_i$$

Two discriminating functions were employed to discriminate more than two clusters, where C_1 is the predicted skin gelatine cluster between porcine and combined bovine and fish gelatine clusters, c_0 is a constant while b_1, b_2 and b_i are regression coefficients for each a AA. Similarly, the C_2 is the predicted cluster of skin gelatine between bovine and fish gelatines.

The F statistics and p-value of the Wilks' Lambda test were computed while individual F-statistics and p-value of significant AAs were determined. Using the training, validation and testing datasets, correct classification of the clusters was examined, and the dissimilarities of porcine, bovine and fish clusters were explored. Value for Fisher distance of the clusters and their p-value were also calculated to verify the cluster dissimilarity.

By utilising the significant AAs ($p < 0.01$) identified by the DAAA, another DA model was carried out and named as DAAAPV. The F statistics, p-value and correct classification of the clusters were compared with the DAAA results, and the best DA model was selected. Based on the correct classification, the best DM model to authenticate the source of skin gelatine was also selected among the PLS-DA, PLS-DAAA, DAAA and DAAAPV.

2.6 Principal component analysis with varimax rotation

A Pearson correlation of principal component analysis (PCA) at α of 0.01 was employed to group skin gelatines into porcine, bovine and fish clusters and identify the AA distribution. The significant AAs ($p < 0.01$) identified by DAAA was transformed and underwent PCA to generate 15 principal components (PCs) known as independent variables. Component score C for y PC number and n sample number can be expressed as:

$$C_{yn} = f_{y1} a_{n1} + f_{y2} a_{n2} + \cdots + f_{yi} a_{ni}$$

Where f is the factor loading (FL), a is the AA concentration, and i is the total number of the AA.

Cumulative variability (CV) of two dimensional PCs entailing PC1 and PC2 were computed for the AA profile exploratory. To clearly group the skin gelatines into porcine, bovine and fish clusters, Varimax rotation at zero (no rotation), two, four and six rotations was carried out. Varimax rotation was stopped at six rotations as the skin gelatine sources had been clearly clustered. The significant AAs contributing to the skin gelatine sources were evaluated.

3.0 Result And Discussion

3.1 Amino acids content in porcine, bovine and fish skin gelatines

This study investigated the distribution of AA content in porcine, bovine and fish skin gelatines. Table 1 shows the amino acid content in each skin gelatine. The presence of 17 AAs in the gelatine was confirmed with a retention time of SS. Glycine was dominant, while His was undetected in the porcine skin gelatine. The ranking of AA concentration in the porcine skin gelatine was as follows: Gly (33.66%) > Pro (12.16%) > Hyp (10.63%) > Ala (9.77%) > Glu (6.54%) > Arg (6.30%) > Lys (3.92%) > Asp (3.48%) > Ser (3.08%) > Leu (2.46%) > Val (2.37%) > Thr (1.79%) > Phe (1.56%) > Ile (1.05%) > Met (0.77%) > Tyr (0.45%) > His (0.00%). In comparison with Hafidz & Yaakob (2011), Azilawati et al. (2015) and Widyaninggar et al. (2012), our study had a similar AA distribution: Gly > Pro, Asp > Ser, and Ile > Met > Tyr. The AA distribution in porcine skin gelatine analysed by a validated and verified method by Abdullah Sani et al. (2021) showed a similar distribution. Although the porcine bone could also be used to produce gelatine, to the authors' knowledge, no report was found on the AA distribution from the porcine bone gelatine.

Based on the ranking of bovine skin gelatine, i.e. Gly (33.83%) > Pro (11.90%) > Hyp (10.89%) > Ala (9.95%) > Glu (6.72%) > Arg (5.95%) > Lys (3.84%) > Asp (3.65%) > Ser (3.11%) > Leu (2.46%) > Val (2.23%) > Thr (1.79%) > Phe (1.47%) > Ile (1.26%) > Met (0.65%) > Tyr (0.28%) > His (0.00%) (Table 1), the AA distribution was similar to the AA distribution of porcine skine gelatine probably due to both of pocine and bovine are mammals. This similarity may render difficulty in differentiating the porcine and bovine skin gelatines. This AA distribution contradicted the finding of Azilawati et al. (2015) and Hafidz & Yaakob (2011) except Gly > Pro, Asp > Ser and Ile > Met > Tyr > His distributions. Valipour et al. (2008) identified the AA distribution of bovine bone gelatine as follows: Gly (17.24%) > Glu (15.56%) > Asp (11.47%) > Pro (9.4%) > Ala (6.67%) > Lys (3.78%) > Thr (3.15%) > Phe (3.15%) > Ser (2.94%) > Arg (2.38%) > Leu (2.27%) > Val (2.09%) > Ile (1.15%) > Met (0.78%) > His (0.67%) > Tyr (0.66%). From this distribution, both bovine skin and bone gelatines had the Gly as the dominant AA, and similar Ile > Met distribution. On contrary, Valipour et al. (2008) identified 0.67% His in the bovine bone gelatine while our result found no His in the bovine skin gelatine.

Table 1 also presents the AA distribution of fish skin gelatine, which followed this ranking: Gly (35.44%) > Ala (9.73%) > Pro (9.43%) > Arg (6.77%) > Glu (6.25%) > Hyp (6.22%) > Ser (6.02%) > Lys (3.81%) > Asp (3.79%) > Thr (2.76%) > Val (2.02%) > Leu (2.02%) > Met (1.81%) > Phe (1.43%) > Ile (1.20%) > His (0.96%) > Tyr (0.33%). The AA distribution of yellowfin tuna (*Thunnus albacares*) skin gelatine was in line with our finding at the Pro > Ar > Glu and Ile > His > Tyr ranking (Nurilmala et al., 2019). Nevertheless, Nawaz et al. (2020) stated that cold-water fish skin gelatine had lower Hyp than the skin gelatine of warm water fish. This finding was supported by a higher Hyp in tilapia (*Oreochromis mossambicus*), yellowfin tuna (*Thunnus albacares*) and blackcarp (*Mylopharyngodon piceus*) than cod (*Gadus morhua*), hake (*Merluccius capensis*) and alaska pollock (*Gadus chalcogrammus*). Of the 17 AAs, our fish skin gelatines had similar AA distribution of Met > Phe > Ile > His > Hyl > Tyr in cod; Gly > Ala > Pro, Glu > Hyp, Thr > Val, and Met > Phe in hake; and : Gly > Ala > Pro and Met > Phe > Ile > His > Hyl > Tyr (Derkach et al., 2020). The bone gelatine of *Ephinephelus sp*. has Gly > Pro > Glu > Ala > Arg > Asp > Leu > Ser > Lys > Thr > Val > Phe > Ile > His > Tyr distribution where only Thr > Val and Phe > Ile > His distribution were similar to our study (Suprayitno, 2019). These findings indicated that each skin and bone gelatine of cold-water fish had their individual AA distribution although some similarities are recorded. Due to this reason, the differention of the skin gelatine of porcin, bovine and fish may possibly be carried out via the statistical analysis.

The ANOVA test in Table 1 shows a significant difference in the mean value of AA among the skin gelatine of porcine, bovine, and fish where skin gelatines with different superscript alphabet were significantly different ($p < 0.01$). The Arg, Pro, Tyr, Met, Val and Ile were significantly different among the three skin gelatines. Specifically, the gelatine of porcine skin had the highest content of Pro, Tyr and Val, and the lowest Ile content. The gelatine of bovine skin had the highest percentage of Ile while had the lowest percentage of Arg, Tyr and Met. The fish skin gelatine had the highest Arg and Met content while it had the lowest Pro and Val content.

The Arg, Pro, Tyr, Met, Val and Ile could be utilised to differentiate porcine and bovine skin gelatines, although the AA distribution between these skin gelatines was similar. However, the content differences between porcine and fish skin gelatines were significant in all AAs except Ala and Lys. Likewise, our study observed a significant difference of all AAs except Asp, Ala, Lys, and Phe in bovine and fish skin gelatines. Nevertheless, the application of the ANOVA test was insufficient to discriminate the three gelatines since more than one AA characterised the gelatines; hence, Abdullah Sani et al. (2021) and (Azilawati et al., 2015) proposed the MDA application to discriminate the skin gelatines.

3.2 Outlier treatment and dataset adequacy

Prior to the MDA, the skin gelatine datasets underwent pre-processing to ensure the dataset fulfilled the MDA prerequisite, including outlier treatment, dataset transformation, and dataset adequacy test (Ismail et al., 2021). The training dataset had 29, 12 and 21 outliers in the porcine, bovine and fish skin gelatines (Table 1), respectively, where our method replaced the outliers with the mean value of each AA (Abdullah Sani et al., 2021). Then, this training dataset was transformed via standardise (n-1) method. Although only negligible reports carried out dataset transformation in their works, our study performed the transformation to fulfil the prerequisite of MDA (Azilawati et al., 2015). Additionally, various dataset transformations are available for MDA, e.g., standardise (n), standard deviation − 1 (n-1), standard deviation − 1 (n), centre, 0 to 1 rescaling, 0 to 100 rescaling, Pareto and log methods (Ismail et al., 2021); however, our study adopted standardise (n-1) as proposed by Abdullah Sani et al. (2021) for gelatine matrix since high AA numbers achieved normality post this transformation.

Table 1 shows the individual KMO value for each AA where Met (0.9274) and Ile (0.6075) had the highest and lowest KMO values, respectively. Comparison of the average KMO value (0.7874) with the guideline from Williams and Brown (2012) study indicated that the dataset adequacy fell on the good ranking (0.7 < KMO < 0.8). Yuswan et al. (2021) and Azilawati et al. (2015) employed MDA without declaring the fulfilment of the dataset adequacy; hence, comparison of the result may not be possible. Nevertheless, other gelatine studies found that KMO > 0.7 signified that the dataset was adequate for MDA (Abdullah Sani et al., 2021). Our KMO value (0.7874) was higher than the gelatine study by Ismail et al. (2021) (KMO = 0.7542). Our KMO value indicated that the dataset was adequate for MDA based on these comparisons.

3.3 Development of model of partial least square discriminant analysis for skin gelatine sources

In this study, the PLS-DA model generated two components to explain the classification ability of the sources of skin gelatine. Table 2 shows two PLS-DA models that provided the classification ability to discriminate the porcine, bovine and fish skin gelatines. The first DM was partial least square-discriminating analysis (PLS-DA) for 17 AAs (PLS-DAAA), while the second DM was PLS-DA for AA with variable importance in the projection (VIP) score > 0.8 (PLS-DAVIPAA). Since too low a cut-off VIP threshold value may lead to a selection of unrelated variables to the gelatine sources, our study followed a VIP score > 0.8 as recommended by Sharin et al. (2021), where AA with a high VIP score could explain most of the variance among the porcine, bovine and fish gelatines. Selection of the best PLS-DA model was made by evaluating the performance of PLS-DAAA and PLS-DAVIPAA models.

The quality of both PLS-DA models was evaluated via R^2Y cumulated (R^2Y cum), R^2X cumulated (R^2X cum) and Q^2 cumulated (Q^2 cum) indices on each component. For DM of PLS-DAVIPAA, the R^2Y cum (0.9356) and R^2X cum (0.8650) were higher than the PLS-DAAA (R^2Y cum = 0.9057 and R^2X cum = 0.7186), indicating that the PLS-DAVIPAA was better in explaining the gelatine clusters and AA contribution to the gelatine clusters, respectively. These findings were associated with the definition of R^2Y cum that is a sum of determination coefficients between the gelatine clusters and two components, while the R2X cum is the sum of determination coefficients between the AA and two components. The R^2Y cum and R^2X cum measured the two components' power to explain the gelatine clusters and AAs. As a generic consequence, the Q^2 cum (0.9320) of PLS-DAVIPAA was also higher than the Q^2 cum (0.8961) of PLS-DAAA, signifying that the two components generated by the PLS-DAVIPAA model had a significant contribution to predictive quality for skin gelatine sources. Additionally, the AAs with VIP score > 0.8 proved to be the main contributors to the predictive quality for skin gelatine sources.

The PLS-DAAA identified 13 significant AAs with descending VIP score, i.e., Tyr (1.4149), Phe (1.3326), Arg (1.2440), Thr (1.0960), Ser (1.0936), Met (1.0924), His (1.0912), Val (1.0783), Gly (1.0783), Hyp (1.0754), Ile (0.9959), Pro (0.9853) and Leu (0.9550) where the Tyr and Leu were the most and least significant AA, respectively. Based on the VIP score, these AAs explain most of the variance among the porcine, bovine and fish gelatines. Hence, these AAs could be used to differentiate the gelatine sources. Also, all 13 AAs of the PLS-DAVIPAA yielded VIP scores > 0.8, confirming the AA significance in discriminating the gelatine sources (Table 2).

Of these AAs, Fig. 1 (a) depicted the 17 AAs plot from PLS-DAAA with individual value of correlation matrix (CMV) for each AA, where His, Hyp, Ser, Thr, Met, Pro and Leu had CMV of 0.95–0.88 and were followed by Val, Arg and Gly with CMV of 0.69–0.63 in component 1. The Ile, Phe, Tyr, Glu, Asp, Lys and Ala had the lowest CMV (0.47–0.064). For component 2, the Tyr and Phe had the highest CMV (0.85–0.83); Arg, Gly, Val, Lys and Ala had the moderate CMV (0.70–0.51); and Glu, Asp, Pro, Leu, Ile, Met, Thr, Ser, His and Hyp had the lowest CMV (0.38–0.04). Figure 1 (b) of PLS-DAVIPAA shows the CMV of 13 AAs where the CMV for each AA in component 1 and 2 had a similar value. Jannat et al. (2018, 2020b) carried out PLS-DA analyses to distinguish porcine, bovine and fish gelatines, but none of them explained the CMV of the detected compounds or amino acids.

Nevertheless, Ismail et al. (2021) classified the AAs into strong (CMV ≥ |0.750|), moderate |0.500| < CMV < |0.749| and weak CMV ≤ |0.499| factor loading for AAs according to the CMV of AAs from principal component analysis (PCA), not PLS-DA. The CMV was used to delineate the AA relationship among them and assign the AA to the gelatine sources (Ismail et al., 2021). Figure 1 (b) exhibits positive correlations based on AA direction proximities; His, Ser, Met and Thr; Gly and Arg; Tyr and Phe; and Leu and Pro. On the contrary, negative correlations of

AAs were observed based on their opposite direction: His, Ser, Met and Thr against Hyp; Tyr and Phe against Ile, Val against Ile, and Leu and Pro against Ile. Arg and Gly did not correlate with Ile since their directions were at 90◦.

Ismail et al. (2021) proposed that the AA's correlations were due to the AA's polarity side chain; however, our study found that only Met has a non-polar side chain although it had a positive correlation with His, Ser and Thr. Further generic grouping of AAs, e.g. basic, carboxylic, hydroxylic and hydrophobic based on the chemical characteristics by Derkach et al. (2020), could not support the AA correlations. The opposite side chains of Gly and Arg and Tyr and Phe also signified that the correlations of AAs were independent of their polarity side chain and generic chemical characteristics. Nevertheless, backbone of the chemical structure may suggest the reason for the positive correlations among the AAs and vice versa. For instance, Met, His, Ser and Thr, and Gly and Arg share HO-CO-CNH_2- backbone; Tyr and Phe share HO-CO-CNH_2-CH_2-benzene backbone while Leu and Pro shares HO-CO- backbone. Furthermore, Hyp and Pro, and Leu and Val were also positioned at close proximity that share HO-CO-pyrrole and HO-CO-NH_2 backbones, repectively.

To assign the AAs to porcine, bovine and fish skin gelatines, the skin gelatines and AA plots shall be overlaid together where the PLS-DA feature of XLSTAT 2019 could not provide in this study. However, the PCA is a preferable method since AA and skin gelatine plots are available in the PCA feature that serves as an exploratory MDA. Hence, in the next section, our study carried out the AAs assignment via PCA.

Table 2 also exhibits the correct classification of PLS-DAAA on the porcine, bovine and fish gelatines. The training and validation datasets exhibited 100% total classification of the porcine, bovine and fish skin gelatines (Table 2), indicating the PLS-DAAA was able to discriminate the gelatines at a 99% confidence level. This result was evident via the small p-value ($p < 0.0001$) of Mahalanobis distance and three distinct skin gelatine clusters, i.e. porcine, bovine and fish, in Fig. 1 (c). Further investigation on the predictive ability of the PLS-DAAA model on the testing dataset showed that it was able to achieve 93.3% of correct classification, where at least 90% skin gelatine of porcine and bovine and 100% of fish skin gelatine were correctly classified. This finding indicated that the PLS-DA with 17 AAs may facilitate the identification of skin gelatine sources, i.e. porcine, bovine and fish, in this study. The PLS-DAVIPAA also exhibited similar correct classification for training and validation datasets. The PLS-DAVIPAA also rendered 93.3% correct classification of its testing dataset, which was similar to the result of PLS-DAAA. Although both PLS-DA models had the same correct classification, Fig. 1 (d) of PLS-DAVIPAA showed that each porcine, bovine and fish skin gelatine were located nearer within their clusters as compared to the skin gelatine plot for PLS-DAAA in Fig. 1 (c). This result was evident in a narrower range of component 2 score of PLS-DAVIPAA than PLS-DAAA, i.e. porcine skin gelatine (-3 to 0), bovine skin gelatines (1 to 4) and fish skin gelatine (-1 to 1) in Fig. 1 (d). On the contrary, the component 2 score of the gelatine sources for PLS-DAAA was as follows: -4 to 0 for porcine skin gelatines, 0 to 5 for bovine skin gelatine, and − 3 to 1 for fish skin gelatine (Fig. 1 (c)). This correct classification also supported the finding on the higher R^2Y cum and Q^2 cum of PLS-DAVIPAA over PLS-DAAA. Hence, this study proposed selecting AAs with VIP scores > 0.8 and employing PLS-DAVIPAA to authenticate skin gelatine sources.

3.4 Development of model of discriminant analysis for skin gelatine sources

Table 3 presents two models of discriminant analysis (DA) i.e. discriminant analysis for 17 AAs (DAAA) and DA entailing AA with p-value < 0.01 (DAAAPV). The DAAA identified 15 AAs that significantly contributed ($p < 0.01$) to the discrimination of porcine, bovine and fish skin gelatines. The decreasing ranking of AAs according to the F-statistic value was as follows: His > Ser > Hyp > Thr > Met > Tyr > Arg > Phe > Pro > Leu > Val > Gly > Ile > Asp > Lys indicating that AA with high F-statistics were more significant than the lower ones in discriminating the sources of skin gelatine. Of the 15 AAs, the DAAAPV selected 13 AAs with a similar ranking of F-statistic value in DAAA, where Asp and Lys were removed from the list since these AAs were the least significant ones.

Figure 2 (a) shows the 17 AA plots of DAAA while Fig. 2 (b) depicts the AA plot of DAAAPV with $p < 0.01$. These figures depicted similar correlations among the AAs with high F-statistic value and $p < 0.01$. Positive AA correlation appeared among His, Ser, Met and Thr; Gly and Arg; Tyr and Phe; and Leu and Pro. There is a negative correlation between His, Ser, Met and Thr versus Hyp; Tyr and Phe versus Ile; Val versus Ile; Leu and Pro versus Ile. Also, Arg and Gly showed no correlation with Val, as evidenced by their 90◦ direction. These correlation patterns were similar to the correlations shown by PLS-DAAA and PLS-DAVIPAA.

The DAAA was able to perform the discrimination role at a 99% confidence level, as shown in 100% classification of the sources of skin gelatines using training and validation datasets. The low $p < 0.0001$ of Fisher distance and three distinct sources of skin gelatines in Fig. 2 (b) supported this finding. Nevertheless, the training dataset had a 96.7% correct classification of the sources of skin gelatines. The DAAAPV had also 100% correctly classified 120 and 40 skin gelatines in the training and testing datasets. On the other hand, the DAAAPV could only classify 30 skin gelatines at 96.7% correct classification, which denoted that both DAAA and DAAPV had similar classification capability.

However, the Fisher distance of DAAA among the clusters, i.e. porcine against bovine and porcine against fish skin clusters, were higher than DAAPV, suggesting that the DAAA could yield more precise classification and lower possibility of cluster overlapping. A more detailed investigation also showed that the DAAA had smaller cluster dispersion in porcine, bovine and fish skin gelatine (Fig. 2 (c)) as compared to DAAAPV in Fig. 2 (d). For DAAA, the porcine cluster had a range of 4.00 for component 1 and 4.19 for component 2; bovine cluster had a

range of 3.42 for component 1 and 4.98 for component 2; and fish cluster had a range of 4.20 for component 1 and 2.53 for component 2. For DAAPV, the porcine cluster had a range of 5.74 for component 1 and 4.03 for component 2; bovine cluster had a range of 3.26 for component 1 and 3.68 for component 2; and fish cluster had a range of 4.81 for component 1 and 4.27 for component 2. These ranges denoted that the DAAA had lower intra-cluster variance that could reduce the possibility of incorrect classification compared to DAAAPV. Hence, the DAAA was the best DA model compared to the DAAAPV. This result signified that no identification of significant AAs via F-statistic value and the p-value is needed if DA is employed to authenticate the source of skin gelatine.

From the results of PLS-DAVIPAA and DAAA, it could be concluded that the DAAA was the best DM for authentication of porcine, bovine and fish skin gelatine sources since the DAAA had 96.7% correct classification of the skin gelatine sources as compared to 93.3% for the PLS-DAVIPAA. This finding was in line with Brereton & Lloyd (2014), which indicated that DA was more suitable than PLS-DA since the number of gelatine samples (observation) were higher than the AA numbers (variables), and there was no missing value in the dataset (Komsta et al., 2018).

3.5 Exploring amino acid profile in skin gelatines

The PCA application in this study aimed to explain the distribution of significantly identified AAs by the DAAA in porcine, bovine and fish skin gelatines. The skin gelatine plots in Fig. 3 (a) – (d) had two principal components (PCs) with cumulative variability (CV) of 78.54% with an eigenvalue (EV) of 3.97 that explained the 13 AAs distribution. However, our study could not achieve these purposes when clusters of different sources were mixed, as shown in Fig. 3 (a). Hair et al. (2014) adopted orthogonal or Varimax rotation and its rotation value since it is superior to other orthogonal rotations, e.g. Equimax and Quartimax, in simplifying the PC structure and providing optimal clusters (Otto, 2017). This study applied Varimax rotation at two, four and six rotation values to enhance variance of factor loadings (FLs) of the PC, reducing dimensionality and facilitating the explanation of 13 AAs distribution in each skin gelatine (de Almeida et al., 2020). Of these Varimax rotations, the four Varimax rotation in Fig. 3 (c) was the optimised rotation where it could reposition all skin gelatines into their clusters.

Figure 3 (e) assigned the AAs to the three clusters by overlaying the skin gelatine and AA plots to investigate their distribution in each skin gelatine. Figure 3 (a) also depicted the absent information in the PLS-DA, such as the dominant, moderate and low AA content in each cluster. The dominant AAs were as follows: Tyr, Phe and Val in porcine gelatine and Met, Thr, Ser, His, Arg and Gly in fish gelatine since these AAs and the clusters were in the same direction. This finding was in line with Abdullah Sani et al. (2021). Our study had a similarity with Azilawati et al. (2015) on Tyr, Met, Thr and Ser in porcine and bovine skin gelatines, respectively. The Pro, Leu and Hyp contents were moderate in both porcine and bovine skin gelatines, while Ile content was moderate in bovine and fish skin gelatines. This moderate content was due to these AAs' direction in the middle of these clusters. Since the Hyp was moderately distributed in porcine and bovine skin gelatines, our result may agree with Yuswan et al. (2021) study that proposed Hyp as one of the biomarkers for halal authentication in gelatine products. The Arg and Gly, and Ile contents were low in porcine and bovine skin gelatines, respectively, since their directions were opposite these clusters. Likewise, fish skin gelatine had low Pro, Leu and Hyp. This finding recommended conducting PCA with four Varimax rotations to (1) ensure all skin gelatines are grouped in their specific clusters and (2) assign the dominant, moderate and low content of AAs in each cluster.

4.0 Conclusion

This study had shown that putatively analysed AAs in skin gelatine via UHPLC-DAD incorporated with DA was able to discriminate the skin gelatine sources. The DA with a higher percentage of correct classification was superior to PLS-DA for discriminating skin gelatine sources. The PCA with four Varimax rotations was also able to assign the skin gelatines to their clusters and provide the AA profile in each cluster. Further study on developing the diagnostic ratio for authentication of skin gelatine sources from this profiling is in the pipeline as a continuation of this study. This study focuses on the classification of skin gelatine sources only since gelatine from this source are the majority being used in gelatine manufacturing in the food industry. Hence, a further study including other gelatine sources such as bone and variations, e.g. blooms, will also be carried out in the near future. Also, this study provided the most common MDA approach, i.e. PLS-DA, DA and PCA, whereas advanced machine learning approaches, i.e. support vector machine, k-means clustering etc., have not been explored. Since this study did not undergo method validation and verification, it may expedite the authentication analysis with less cost and time. Based on this study, the authority may adopt and regulate a standard ad-hoc test to authenticate skin gelatine products.

Declarations

5.0 Acknowledgement

The authors would like to acknowledge the Ministry of Higher Education Malaysia for granting Konsortium Institut Halal IPT Malaysia (KIHIM) a research grant 63900911-10205 and Malaysia Halal Analysis Centre (MyHAC) of the Department of Islamic Development Malaysia for the funding, laboratory's facilities, and assistance.

6.0 Conflict of interest statement

We declare no conflict of interest.

7.0 Research involving human participants and/or animals

We declare no human and/or animals involved in this study.

8.0 Informed consent and credit author statement

All authors have granted permission in full knowledge for the publication of this study. The credit author statement as follows: Azilawati Mohd Ismail for methodology; Muhamad Shirwan Abdullah Sani for conceptualisation, data curation and writing of original draft; Azman Azid for software and validation; and Mohd Saiful Samsudin for visualisation and investigation.

References

1. Abdullah Sani, M. S., Ismail, A. M., Azid, A., & Samsudin, M. S. (2021). Establishing forensic food models for authentication and quantification of porcine adulterant in gelatine and marshmallow. Food Control, *130*(June), 108350. https://doi.org/10.1016/j.foodcont.2021.108350
2. Andrada, M. F., Vega-Hissi, E. G., Estrada, M. R., & Garro Martinez, J. C. (2015). Application of k-means clustering, linear discriminant analysis and multivariate linear regression for the development of a predictive QSAR model on 5-lipoxygenase inhibitors. Chemometrics and Intelligent Laboratory Systems, *143*, 122–129. https://doi.org/10.1016/j.chemolab.2015.03.001
3. Azilawati, M. I., Hashim, D. M., Jamilah, B., & Amin, I. (2015). RP-HPLC method using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate incorporated with normalisation technique in principal component analysis to differentiate the bovine, porcine and fish gelatins. Food Chemistry, *172*, 368–376. https://doi.org/10.1016/j.foodchem.2014.09.093
4. Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. Journal of Chemometrics, *28*(4), 213–225. https://doi.org/10.1002/cem.2609
5. de Almeida, F. A., Leite, R. R., Gomes, G. F., Gomes, J. H. de F., & de Paiva, A. P. (2020). Multivariate data quality assessment based on rotated factor scores and confidence ellipsoids. Decision Support Systems, *129*(October), 113173. https://doi.org/10.1016/j.dss.2019.113173
6. Department of Standards Malaysia. (2019). *Malaysian Standard MS 1500: 2019 Halal food - General requirements (Third revision)*.
7. Derkach, S. R., Voron, N. G., Kuchina, Y. A., & Kolotova, D. S. (2020). *Modified Fish Gelatin as an Alternative to Mammalian Gelatin in Modern Food Technologies*. 27–29.
8. Hafidz, R., & Yaakob, C. (2011). Chemical and functional properties of bovine and porcine skin gelatin. International Food Research Journal, *18*, 813–817. http://ifrj.upm.edu.my/18 (02) 2011/(48) IFRJ-2010-159.pdf
9. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate data analysis. In *Pearson Education Limited*. https://doi.org/10.1007/978-3-319-01517-0_3
10. Idris, M. H. H., Manaf, Y. N., Desa, M. N. M., Hashim, A. M., Sani, M. S. A., Zaki, N. N. M., Yuswan, M. H., Kamaruddin, M. S., Hassan, M. S., Yusof, Y. A., & Mustafa, S. (2021). A conjunction of sn-2 fatty acids and overall fatty acid composition combined with chemometric techniques increase the effectiveness of lard detection in fish feed. Chemometrics and Intelligent Laboratory Systems, *213*(March), 104308. https://doi.org/10.1016/j.chemolab.2021.104308
11. Ismail, A. M., Sani, M. S. A., Azid, A., Zaki, N. N. M., Arshad, S., Tukiran, N. A., Abidin, S. A. S. Z., Samsudin, M. S., & Ismail, A. (2021). Food forensics on gelatine source via ultra–high–performance liquid chromatography diode–array detector and principal component analysis. SN Applied Sciences, *3*(79), 1–19. https://doi.org/10.1007/s42452-020-04061-7
12. Jannat, B., Ghorbani, K., Kouchaki, S., Sadeghi, N., Eslamifarsani, E., Rabbani, F., & Beyramysoltan, S. (2020). Distinguishing tissue origin of bovine gelatin in processed products using LC/MS technique in combination with chemometrics tools. *Food Chemistry*, *319*(September 2018). https://doi.org/10.1016/j.foodchem.2020.126302
13. Jannat, B., Ghorbani, K., Shafieyan, H., Kouchaki, S., Behfar, A., Sadeghi, N., Beyramysoltan, S., Rabbani, F., Dashtifard, S., & Sadeghi, M. (2018). Gelatin speciation using real-time PCR and analysis of mass spectrometry-based proteomics datasets. Food Control, *87*, 79–87. https://doi.org/10.1016/j.foodcont.2017.12.006
14. Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, *39*(1), 31–36.

15. Komsta, L., Heyden, Y. Vander, & Sherma, J. (2018). Chemometrics in chromatography. In *Taylor & Francis Group, LLC*.
16. Mutalib, S. A., Muin, N. M., Abdullah, A., Hassan, O., Wan Mustapha, W. A., Abdullah Sani, N., & Maskat, M. Y. (2015). Sensitivity of polymerase chain reaction (PCR)-southern hybridisation and conventional PCR analysis for Halal authentication of gelatin capsules. LWT - Food Science and Technology, *63*(1), 714–719. https://doi.org/10.1016/j.lwt.2015.03.006
17. Nawaz, A., Li, E., Irshad, S., Xiong, Z., Xiong, H., Shahbaz, H. M., & Siddique, F. (2020). Valorisation of fisheries by-products: Challenges and technical concerns to food industry. Trends in Food Science and Technology, *99*, 34–43. https://doi.org/10.1016/j.tifs.2020.02.022
18. Nurilmala, M., Pertiwi, R. M., Nurhayati, T., Fauzi, S., Batubara, I., & Ochiai, Y. (2019). Characterisation of collagen and its hydrolysate from yellowfin tuna Thunnus albacares skin and their potencies as antioxidant and antiglycation agents. Fisheries Science, *85*(3), 591–599. https://doi.org/10.1007/s12562-019-01303-5
19. Otto, M. (2017). Chemometrics: Statistics and computer application in analytical chemistry. In Wiley-VCH Verlag GmbH & Co. https://doi.org/10.1007/s00216-017-0503-7
20. Rohman, A., Windarsih, A., Erwanto, Y., & Zakaria, Z. (2020). Review on analytical methods for analysis of porcine gelatine in food and pharmaceutical products for halal authentication. Trends in Food Science and Technology, *101*, 122–132. https://doi.org/10.1016/j.tifs.2020.05.008
21. Romi Mukherjee, S. (2014). Global Halal: Meat, money, and religion. Religions, *5*(1), 22–75. https://doi.org/10.3390/rel5010022
22. Saiful, M., Azid, A., Iskandar, S., Shirwan, M., Sani, A., & Lananan, F. (2019). Comparison of prediction model using spatial discriminant analysis for marine water quality index in mangrove estuarine zones. *Marine Pollution Bulletin*, *141*(February 2018), 472–481. https://doi.org/10.1016/j.marpolbul.2019.02.045
23. Sani, M. S. A., Yuswan, M. H., Desa, M. N. M., & Azid, A. (2020). Procedure of method validation and verification, and multivariate data analysis of chemical measurement for halal-testing laboratories. *Scientific Conference World Halal Summit 2019*, *1*, 1–13.
24. Sharin, S. N., Sani, M. S. A., Jaafar, M. A., Yuswan, M. H., Kassim, N. K., Manaf, Y. N., Wasoh, H., Zaki, N. N. M., & Hashim, A. M. (2021). Discrimination of Malaysian stingless bee honey from different entomological origins based on physicochemical properties and volatile compound profiles using chemometrics and machine learning. Food Chemistry, *346*(June 2020), 128654. https://doi.org/10.1016/j.foodchem.2020.128654
25. Suprayitno, E. (2019). Amino acid marshmallow profile from grouper bone gelatin. International Journal of Scientific and Research Publications, *9*(11), p9512. https://doi.org/10.29322/ijsrp.9.11.2019.p9512
26. Valipour, P., Talebian, A., & Ebadi, A. G. (2008). Gelatin preparation from cow's bone by heat-pressure method. Asian Journal of Chemistry, *20*(2), 1337–1342.
27. Widyaninggar, A., Triwahyudi, Triyana, K., & Rohman, A. (2012). Differentiation between porcine and bovine gelatin in commercial capsule shells based on amino acid profiles and principal component analysis. Indonesian Journal of Pharmacy, *23*(2), 104–109.
28. Williams, B., & Brown, T. (2012). Exploratory factor analysis: A five-step guide for novices. Journal of Emergengy Primary Health Care, *8*(3).
29. Yap, B. K., & Gam, L. H. (2019). Differentiation of bovine from porcine gelatin capsules using gel electrophoresis method. *Food Chemistry*, *274*(August 2018), 16–19. https://doi.org/10.1016/j.foodchem.2018.08.111
30. Yuswan, M. H., Nurul, N. H., Mohamad, H., Keso, S., Mohamad, N. A., Tengku, T. S., Ismail, N. F., Abdul Manaf, Y. N., Mohd Hashim, A., Mohd Desa, M. N., Yusof, Y. A., & Mustafa, S. (2021). Hydroxyproline determination for initial detection of halal-critical food ingredients (gelatin and collagen). *Food Chemistry*, *337*(April 2020), 1–7. https://doi.org/10.1016/j.foodchem.2020.127762

Tables

Table 1: Amino acid percentage, number of outliers, and sampling adequacy result of porcine, bovine and fish skin gelatines

Amino acid	Retention time, min[1]	Amino acid concentration in testing dataset, %[1,2]			Number of outliers in testing dataset[1,3,4]			Sampling adequacy value[1,5]
		Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	
L-Hydroxyproline (Hyp)	1.774	10.63 ± 0.45^{a}	10.89 ± 0.67^{a}	6.22 ± 0.19^{b}	4 (P3, P18, P19 and P40)	1 (B24)	1 (F37)	0.9048
L-Histidine (His)	1.887	0.00 ± 0.01^{b}	0.00 ± 0.00^{b}	0.96 ± 0.05^{a}	3 (P18, P19, P21, P28 and P40)	1 (B29)	0	0.8519
L-Serine (Ser)	2.543	3.08 ± 0.15^{b}	3.12 ± 0.15^{b}	6.02 ± 0.16^{a}	1 (P40)	1 (B24)	1 (F9)	0.7196
L-Arginine (Arg)	2.610	6.30 ± 0.29^{b}	5.95 ± 0.25^{c}	6.77 ± 0.20^{a}	1 (P9)	1 (B24)	1 (F37)	0.6317
Glycine (Gly)	2.776	33.66 ± 1.05^{b}	33.83 ± 0.93^{b}	35.44 ± 0.88^{a}	1 (P9)	1 (B24)	1 (F9 and F39)	0.7586
L-Aspartic acid (Asp)	2.987	3.48 ± 0.36^{b}	3.65 ± 0.34^{ab}	3.79 ± 0.28^{a}	2 (P4, P15)	0	1 (F13 and F30)	0.8428
L-Glutamic acid (Glu)	3.365	6.54 ± 0.48^{a}	6.72 ± 0.47^{a}	6.25 ± 0.33^{b}	1 (P15)	0	1 (F13 and F30)	0.8320
L-Threonine (Thr)	3.720	1.79 ± 0.08^{b}	1.79 ± 0.07^{b}	2.76 ± 0.08^{a}	1 (P9)	1 (B24)	1 (F9)	0.7721
L-Alanine (Ala)	4.077	9.77 ± 0.51^{a}	9.95 ± 0.57^{a}	9.73 ± 0.33^{a}	1 (P15)	0	2 (F13 and F30)	0.7139
L-Proline (Pro)	4.668	12.16 ± 0.32^{a}	11.90 ± 0.36^{b}	9.43 ± 0.18^{c}	2 (P9, P15)	1 (B13)	2 (F13 and F39)	0.8938
L-Lysine (Lys)	6.009	3.92 ± 0.31^{a}	3.84 ± 0.40^{a}	3.81 ± 0.20^{a}	1 (P15)	0	1 (F13 and F30)	0.7901
L-Tyrosine (Tyr)	6.107	0.45 ± 0.03^{a}	0.28 ± 0.03^{c}	0.33 ± 0.03^{b}	2 (P6 and P40)	0	1 (F38)	0.8049
L-Methionine (Met)	6.351	0.77 ± 0.04^{b}	0.65 ± 0.18^{c}	1.81 ± 0.08^{a}	1 (P8)	0	3 (F9, F37 and F38)	0.9274
L-Valine (Val)	6.624	2.37 ± 0.04^{a}	2.23 ± 0.06^{b}	2.02 ± 0.04^{c}	2 (P9 and P15)	1 (B13)	1 (F13)	0.7890
L-Isoleucine (Ile)	7.940	1.05 ± 0.03^{c}	1.26 ± 0.03^{a}	1.20 ± 0.04^{b}	2 (P8 and P15)	1 (B5)	0	0.6075
L-Leucine (Leu)	8.052	2.46 ± 0.04^{a}	2.46 ± 0.07^{a}	2.02 ± 0.06^{b}	2 (P9 and P15)	1 (B13)	1 (F13)	0.7797
L-Phenylalanine (Phe)	8.184	1.56 ± 0.05^{a}	1.47 ± 0.10^{b}	1.43 ± 0.05^{b}	2 (P10 and P11)	2 (B12 and B24)	3 (F8, F26 and F37)	0.6588
Total outliers	nr	nr	nr	nr	29	12	21	nr
Average KMO value	nr	nr	nr	nr	nr	nr	nr	0.7874

Note: [1]nr = not related.

[2]Different superscript alphabets indicated a significant difference in average relative error mean.

[3]Number of detected outliers by Grubbs and Dixon tests.

[4]Skin gelatine in parenthesis indicates the outlier presence.

[5]Sampling adequacy test by Kaiser-Meyer-Olkin (KMO) test.

Table 2: Classification matrix of training, validation and testing datasets of partial least square – discriminant analysis

Discriminating model (DM)	Discriminating model quality		Dataset	Correct classification, %	Number of gelatines, Mahalanobis distance value and p-value of Mahalanobis distance in skin gelatine cluster[3,4]			Total skin gelatine
	R^2Y cum, R^2X cum and Q^2 cum	Ranking of significant amino acid ($p < 0.01$)[1,2]			Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	
Partial least square – discriminant analysis for 17 amino acids (PLS-DAAA)	R^2Y cum: 0.9057 R^2X cum: 0.7186 Q^2 cum: 0.8961	Tyr (1.4149) > Phe (1.3326) > Arg (1.2440) > Thr (1.0960) > Ser (1.0936) > Met (1.0924) > His (1.0912) > Val (1.0783) > Gly (1.0783) > Hyp (1.0754) > Ile (0.9959) > Pro (0.9853) > Leu (0.9550)	Training dataset					
			Porcine gelatine	100	40 (0, 1)	0 (290, < 0.0001)	0 (2721, < 0.0001)	40
			Bovine gelatine	100	0 (290, < 0.0001)	40 (0, 1)	0 (2720, < 0.0001)	40
			Fish gelatine	100	0 (2721, < 0.0001)	0 (2720, < 0.0001)	40 (0, 1)	40
			Total	100	40	40	40	120
			Validation dataset					
			Porcine gelatine	100	16 (0, < 1)	0 (290, < 0.0001)	0 (290, < 0.0001)	16
			Bovine gelatine	100	0 (290, < 0.0001)	10 (0, 1)	0 (0, 1)	10
			Fish gelatine	100	0 (2721, < 0.0001)	0 (2720, < 0.0001)	14 (2720, < 0.0001)	14
			Total	100	16	10	14	40
			Testing dataset					
			Porcine gelatine	90	9 (0, < 1)	1 (290, < 0.0001)	0 (290, < 0.0001)	10
			Bovine gelatine	90	1 (290, < 0.0001)	9 (0, 1)	0 (0, 1)	10
			Fish gelatine	100	0 (2721, < 0.0001)	0 (2720, < 0.0001)	10 (2720, < 0.0001)	10
			Total	93.3	10	10	10	30
PLS-DA for AA with variable importance in the projection (VIP) > 0.8 (PLS-DAVIPAA)	R^2Y cum: 0.9356 R^2X cum: 0.8650 Q^2 cum: 0.9320 -	Tyr (1.2776) > Phe (1.1688) > Arg (1.1315) > Gly (0.9944) > Val (0.9753) > Thr (0.9674) > Ser (0.9618) > Met (0.9566) >	Training dataset					
			Porcine gelatine	100	40 (0, < 1)	0 (290, < 0.0001)	0 (2721, < 0.0001)	40
			Bovine gelatine	100	0 (290, < 0.0001)	40 (0, < 1)	0 (2720, < 0.0001)	40
			Fish gelatine	100	0 (2721, < 0.0001)	0 (2720, < 0.0001)	40 (0, < 1)	40
			Total	100	40	40	40	120

His (0.9559) > Hyp (0.9404) > Pro (0.8697) > Ile (0.8578) > Leu (0.8457) -					
Validation dataset					
Porcine gelatine	100	11 (0, < 1)	0 (290, < 0.0001)	0 (2721, < 0.0001)	11
Bovine gelatine	100	0 (290, < 0.0001)	15 (0, < 1)	0 (2720, < 0.0001)	15
Fish gelatine	100	0 (2721, < 0.0001)	0 (2720, < 0.0001)	14 (0, < 1)	14
Total	100	11	15	14	40
Testing dataset					
Porcine gelatine	90	9 (0, < 1)	1 (290, < 0.0001)	0 (2721, < 0.0001)	10
Bovine gelatine	90	1 (290, < 0.0001)	9 (0, < 1)	0 (2720, < 0.0001)	10
Fish gelatine	100	0 (2721, < 0.0001)	0 (2720, < 0.0001)	10 (0, < 1)	10
Total	93.3	10	10	10	30

Note: [1]Value in parenthesis was an F-statistic value of significant amino acid with variable importance in the projection (VIP) > 0.8.

[2]Hyp = L-Hydroxyproline, His = L-Histidine, Ser = L-Serine, Arg = L-Arginine, Gly = Glycine, Asp = L-Aspartic acid, Glu = L-Glutamic acid, Thr = L-Threonine,

Ala = L-Alanine, Pro = L-Proline, Lys = L-Lysine, Tyr = L-Tyrosine, Met = L-Methionine, Val = L-Valine, Ile = L-Isoleucine, Leu = L-Leucine and Phe = L-Phenylalanine.

[3]Values in parenthesis were Fisher distance value and p-value of Fisher distance, respectively.

[4]Calculated p-value of Fisher distance < 0.01 indicated three clusters were significantly different.

Table 3: Classification matrix of training, validation, and testing datasets of discriminant analysis

Discriminating model (DM)	Discriminating model quality				Dataset	Correct classification, %	Number of gelatines, Fisher distance value and p-value of Fisher distance in skin gelatine cluster [2,3]			Total skin gelatine
	Wilks' Lambda test	Ranking of significant AA ($p < 0.01$)					Porcine skin gelatine	Bovine skin gelatine	Fish skin gelatine	
		AA[1]	F statistics	p-value						
Discriminant analysis (DA) for 17 amino acids (DAAA)	F statistics = 1032.51 p-value = < 0.0001	His	10974.9314	< 0.0001	Training dataset					
		Ser	4618.1344	< 0.0001	Porcine gelatine	100	40 (0, 1)	0 (188, < 0.0001)	0 (2277, < 0.0001)	40
		Hyp	3791.2488	< 0.0001	Bovine gelatine	100	0 (188, < 0.0001)	40 (0, 1)	0 (2202, < 0.0001)	40
		Thr	2153.3171	< 0.0001	Fish gelatine	100	0 (2277, < 0.0001)	0 (2202 < 0.0001)	40 (0,1)	40
		Met	1251.6230	< 0.0001	Total	100	40	40	40	120
		Tyr	409.6914	< 0.0001	Validation dataset					
		Arg	269.0799	< 0.0001	Porcine gelatine	100	16 (0, 1)	0 (188, < 0.0001)	0 (2277, < 0.0001)	16
		Phe	219.2897	< 0.0001	Bovine gelatine	100	0 (188, < 0.0001)	12 (0, 1)	0 (2202, < 0.0001)	12
		Pro	187.2098	< 0.0001	Fish gelatine	100	0 (2277, < 0.0001)	0 (2202 < 0.0001)	12 (0,1)	12
		Leu	153.1016	< 0.0001	Total	100	16	12	12	40
		Val	119.4813	< 0.0001	Testing dataset					
		Gly	102.5208	< 0.0001	Porcine gelatine	100	10 (0, 1)	0 (188, < 0.0001)	0 (2277, < 0.0001)	10
		Ile	53.5884	< 0.0001	Bovine gelatine	90	1 (188, < 0.0001)	9 (0, 1)	0 (2202, < 0.0001)	10
		Asp	6.6400	0.0019	Fish gelatine	100	0 (2277, < 0.0001)	0 (2202 < 0.0001)	10 (0,1)	10
		Lys	5.6578	0.0045	Total	96.7	11	9	10	30
DA for AA with p-value < 0.01 (DAAAPV)	F statistics = 668.63 p-value = < 0.0001	His	7136.2174	< 0.0001	Training dataset					
		Ser	2984.0792	< 0.0001	Porcine gelatine	100	40 (0, 1)	0 (162, < 0.0001)	0 (2017, < 0.0001)	40
		Hyp	2492.4176	< 0.0001	Bovine gelatine	100	0 (162, < 0.0001)	40 (0, 1)	0 (2298, < 0.0001)	40
		Thr	1565.7871	< 0.0001	Fish gelatine	100	0 (2017, <	0 (2298, <	40 (0, 1)	40

					0.0001)	0.0001)		
Met	1014.0161	< 0.0001	Total	100	40	40	40	120
Tyr	279.6010	< 0.0001	Validation dataset					
Arg	193.7758	< 0.0001	Porcine gelatine	100	16 (0, 1)	0 (162, < 0.0001)	0 (2017, < 0.0001)	16
Phe	146.9638	< 0.0001	Bovine gelatine	100	0 (162, < 0.0001)	12 (0, 1)	0 (2298, < 0.0001)	12
Pro	130.1654	< 0.0001	Fish gelatine	100	0 (2017, < 0.0001)	0 (2298, < 0.0001)	12 (0, 1)	12
Leu	104.0466	< 0.0001	Total	100	16	12	12	40
Val	84.9554	< 0.0001	Testing dataset					
Gly	66.5813	< 0.0001	Porcine gelatine	100	10 (0, 1)	0 (162, < 0.0001)	0 (< 0.0001)	10
Ile	33.8023	< 0.0001	Bovine gelatine	90	1 (162, < 0.0001)	9 (0, 1)	0 (2017, < 0.0001)	10
			Fish gelatine	100	0 (2017, < 0.0001)	0 (2298, < 0.0001)	10 (2298, < 0.0001)	10
			Total	96.7	11	9	10 (0, 1)	30

Note: [1]Hyp = L-Hydroxyproline, His = L-Histidine, Ser = L-Serine, Arg = L-Arginine, Gly = Glycine, Asp = L-Aspartic acid, Glu = L-Glutamic acid, Thr = L-Threonine,

Ala = L-Alanine, Pro = L-Proline, Lys = L-Lysine, Tyr = L-Tyrosine, Met = L-Methionine, Val = L-Valine, Ile = L-Isoleucine, Leu = L-Leucine and Phe = L-Phenylalanine.

[2]Value in parenthesis were Fisher distance value and p-value of Fisher distance, respectively.

[3]Calculated p-value of Fisher distance < 0.01 indicated three clusters were significantly different.

Figures

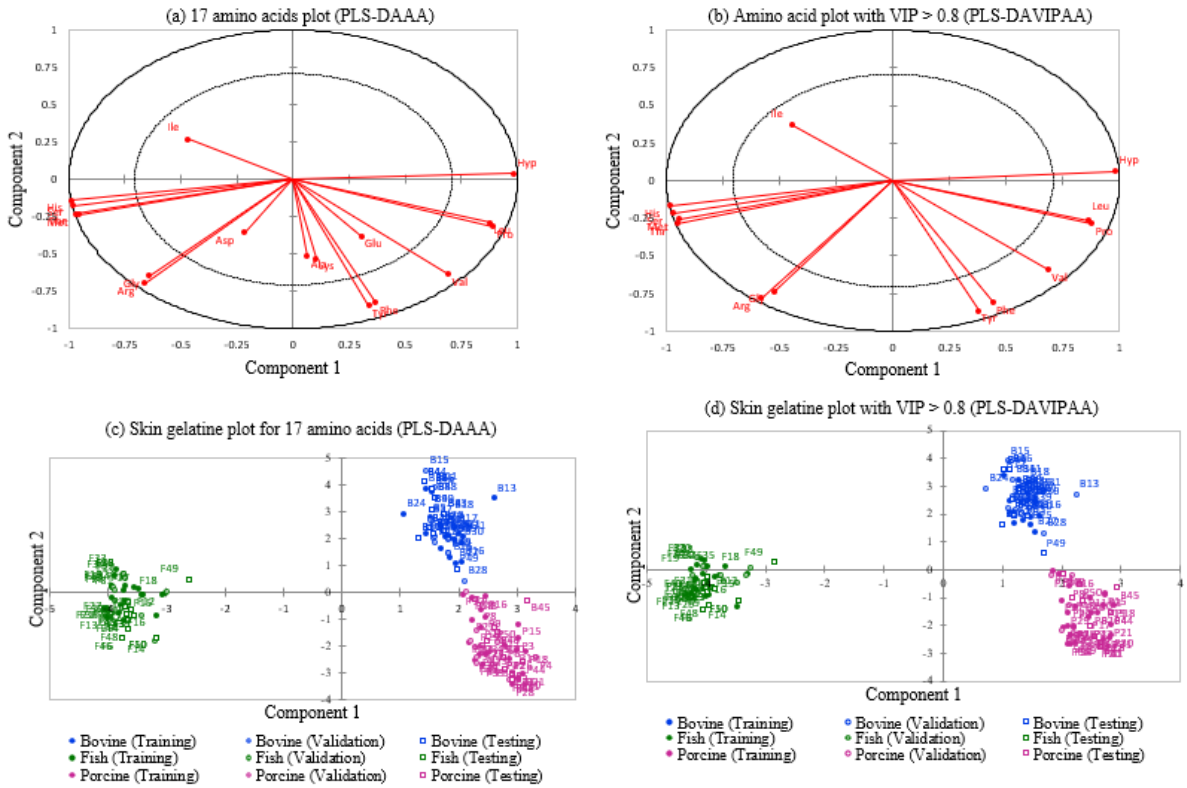

Figure 1: (a) 17 amino acids plot (PLS-DAAA), (b) amino acid plot with VIP > 0.8 (PLS-DAVIPAA), (c) skin gelatine plot for 17 amino acids (PLS-DAAA) and (d) skin gelatine plot with VIP > 0.8 (PLS-DAVIPAA) via partial least square-discriminant analysis

Figure 1

(a) 17 amino acids plot (PLS-DAAA), (b) amino acid plot with VIP > 0.8 (PLS-DAVIPAA), (c) skin gelatine plot for 17 amino acids (PLS-DAAA) and (d) skin gelatine plot with VIP > 0.8 (PLS-DAVIPAA) via partial least square-discriminant analysis

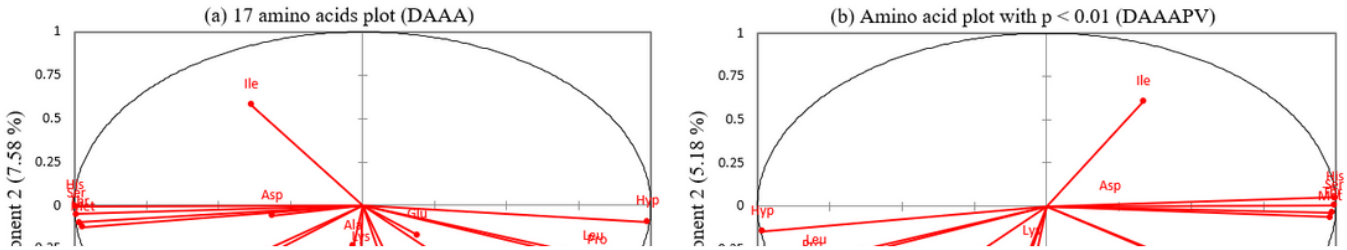

Figure 2

(a) 17 amino acids plot (DAAA), (b) skin gelatine plot for 17 amino acids (DAAA), (c) amino acid plot with p < 0.01 (DAAAPV) and (d) skin gelatine plot with p < 0.01 via discriminant analysis

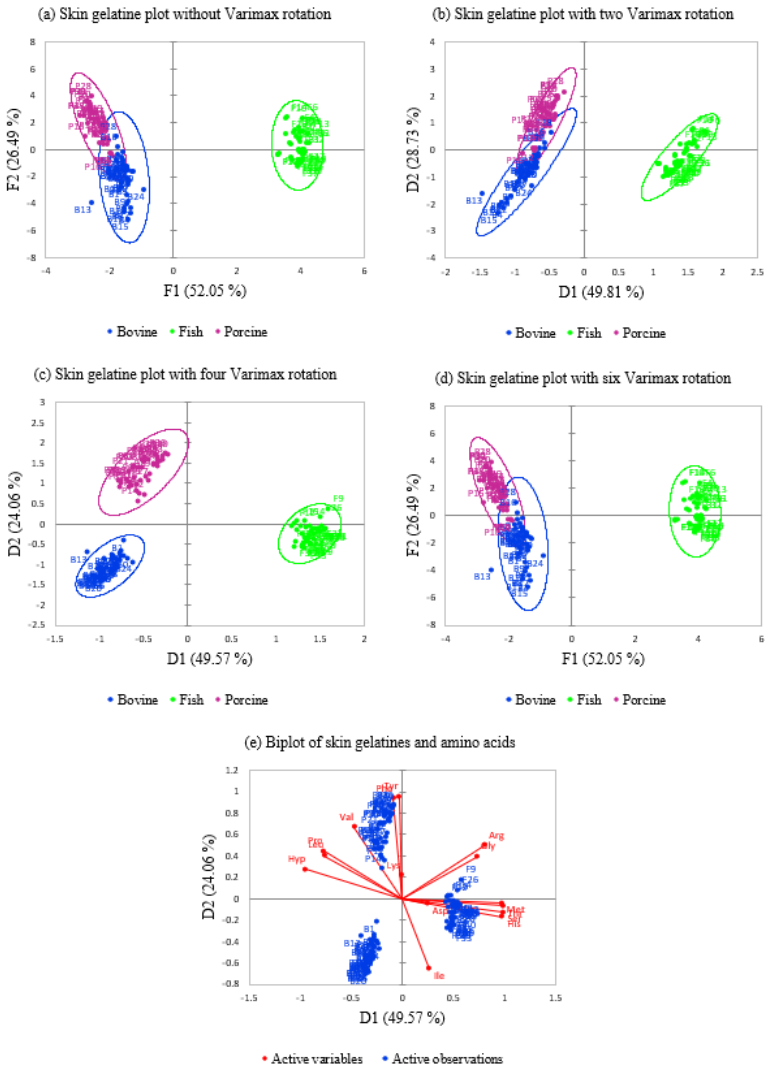

Figure 3: Skin gelatine plots with (a) no Varimax rotation, (b) two Varimax rotation, (c) four Varimax rotation and (d) six Varimax rotation, and (e) biplot of skin gelatines and amino acids with four Varimax rotation

Figure 3

Skin gelatine plots with (a) no Varimax rotation, (b) two Varimax rotation, (c) four Varimax rotation and (d) six Varimax rotation, and (e) biplot of skin gelatines and amino acids with four Varimax rotation